Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 9, No. 1, February 2024, Pp. 29-44

Mental health prediction model on social media data using CNN-BiLSTM

Abdurrahim[1], Dhomas Hatta Fudholi*[1]
Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia, Indonesia[1]

Article Info

Keywords:
Mental Health, Natural Language Processing, Text Analytic, Classification, CNN-BiLSTM

Article history:
Received: September 08, 2023
Accepted: December 17, 2023
Published: February 28, 2024

Cite:
Abdurrahim and Dhomas Hatta Fudholi, "Mental Health Prediction Model on Social Media Data Using CNN-BiLSTM", KINETIK, vol. 9, no. 1, Feb. 2024. Retrieved from
https://kinetik.umm.ac.id/index.php/kinetik/article/view/1849

*Corresponding author.
Dhomas Hatta Fudholi
E-mail address:
hatta.fudholi@uii.ac.id

Abstract

Social media has transformed into a global platform for expression and interaction where users can share photos, images, and videos. The rapid development and widespread use of social media afford the opportunity to analyze the construction of social life in societies and communities. As a result of alterations in lifestyle during the COVID-19 pandemic, mental health disorders increased. Mental health is a complex disease involving numerous individual, socioeconomic, and clinical variables. Natural language processing and analysis methods are required to address this complexity. The classification of mental health-related texts, which can serve as early warnings and early diagnoses, is facilitated by analytical and natural language processing techniques. In this investigation, a CNN-BiLSTM model was utilized, which was aided by a FastText-based word weighting method. The utilized data set consists of texts on mental health with labels such as borderline personality disorder (BPD), anxiety, depression, bipolar, mentalillness, schizophrenia, and poison. There are 35000 training records and 6108 test records. The data will undergo a data cleansing procedure, which will include lower text stages, number removal, reading mark removal, and stopword removal. Modeling with CNN-BiLSTM and FastText weighting yielded an F1-Score and accuracy of 85% and 85%, respectively. In comparison to the Bi-LSTM model, the F1-Score and accuracy were both 83%.

1. Introduction

Mental disorders have a significant impact on a substantial portion of the global population, exceeding 1 billion individuals. This corresponds to a prevalence rate of approximately 1,100,075,000 people, which accounts for approximately 16% of the world's population. This condition exhibits a similar prevalence among both genders, with around 537,698,000 cases reported in women and 572,376,000 cases reported in men [1]. According to a study conducted in 2018 on the general health status of the Indonesian population, it was found that a significant proportion of individuals aged 15 and above, specifically 19 million, experience emotional and mental problems. Additionally, the study revealed that 12 million individuals within the same age group are afflicted with depression. According to the data, the incidence rate of individuals experiencing a disorder is approximately 20% of the population, indicating a probable prevalence of mental health issues among 1 in 5 individuals [2]. The significance of mental health is on par with that of physical health. The domain of mental health encompasses the potential impact on cognitive processes, affective states, and behavioral patterns [3]. There are several types of mental health illnesses, including but not limited to depression, suicidal thoughts, bipolar disorder, autism spectrum disorder (ASD), anxiety disorder, and schizophrenia, which can exert detrimental impacts on an individual's physical well-being [4].

The text consists of a variety of representations that reflect an individual's overt emotional state, such as posts on social media platforms, interview transcriptions, and clinical data incorporated within the patient's account of their mental well-being [4]. In recent years, there has been a significant utilization of Natural Language Processing techniques in the domains of health and biomedicine [5], [6]. Natural Language Processing is a subfield within a field of artificial intelligence (AI) that facilitates the examination of extensive textual data for many purposes, including but not limited to information extraction and sentiment analysis [7]. Natural Language Processing has the potential to be utilized in the identification of emotions and the provision of mental health surveillance [8]–[10]. The identification of mental health conditions can be classified as jobs involving text categorization. The aforementioned measures are implemented in order to facilitate early detection, implement preventive strategies, and offer direction for therapeutic interventions.

Numerous research studies have been conducted to identify mental health conditions, with particular emphasis on distinct domains such as suicide [11]–[13] and depression [14]–[16]. The data sources included in this study are various, encompassing specific social media platforms [17]–[19], as well as non-clinical text sources [20]. The study utilized the mental health corpus and mental health Reddit data in order make predictions about mental health [21], [22].

Deep learning techniques, namely neural networks like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM), have demonstrated high levels of accuracy in several applications [23]–[26]. CNNs employ convolutional techniques to capture local aspects of text within the domain of natural language processing (NLP). However, they encounter challenges when attempting to account for long-range dependencies within textual data and text sequences [27]. The Long Short-Term Memory (LSTM) model is limited to utilizing solely the information present in the forward text, without incorporating any information from the backward context. Hence, this study proposes the utilization of bi-directional long-term memory (Bi-LSTM) networks, which enable the simultaneous access of both forward and backward information. This approach enhances model prediction and effectively incorporates contextual information into the model. However, the computation required for large Bi-LSTM input dimensions can result in substantial overhead. Hence, the combination of CNN and Bi-LSTM was achieved by using a CNN layer to create a pooling layer, which was subsequently fed into the LSTM layer. The proposed methodology has the potential to decrease the vector dimensionality of the word matrix derived from the initial dataset, while utilizing the Bi-LSTM model for the purpose of classification [25].

Many studies have employed the Convolutional Neural Network-Bidirectional Long Short-Term Memory (CNN-BiLSTM) architecture in conjunction with datasets pertaining to mental health. Based on the findings of the study conducted by researchers [28] it was observed that the Multi-Head Attention with Bidirectional Long Short Term Memory and Convolutional Neural Network model had a high level of accuracy, specifically reaching 97.8%. The dataset utilized in these research investigations encompasses the interactions between patients and clinicians sourced from Webmd and Healthtap. It encompasses a total of 2086 data points from Webmd and 5328 data points from Healthtap. In a previous investigation [29] researchers utilized datasets obtained from Reddit and Twitter to examine the performance of LSTM, Bi-LSTM, and BERT+KD models in classifying depression and anxiety labels. The Reddit dataset consisted of 75,000 depression labels and 80,000 anxiety labels, while the Twitter dataset included 25 depression labels and 15,000 anxiety labels. The experimental results indicated accuracy levels of 94%, 96%, and 98% for the LSTM, Bi-LSTM, and BERT+KD models, respectively. The study conducted by Trotzek et al. [30] centered on the timely identification of depression using fasttext features, glove embedding, and CNN models. The outcome of the study yielded an F1-Score of 73%. In a separate investigation [31] the utilization of fasttext techniques and XGBoost models was implemented on datasets related to depression and suicidal risk, resulting in an accuracy rate of 78%.

Researchers [21] employed a multi-class classification method to examine the Reddit mental health dataset in a connected investigation. The dataset comprised of the titles and post texts from 17,159 posts across 13 subreddits, categorized into six distinct groups: r/depression, r/anxiety, r/ptsd, r/adhd, r/bipolar, and none. The study utilized deep learning models such as LSTM, BERT, and RoBERTa. The LSTM model attained 76% accuracy rate in assessing post+titles data, but the BERT and RoBERTa models achieved 87% and 89% accuracy, respectively. A different research [22] using a dataset from Reddit that had five distinct class labels: ADHD, anxiety, bipolar disorder, depression, PTSD, and none. The dataset utilized in this study comprised 16,930 reddit posts, which were segregated into three subsets: training (13,726 posts), development (1,716 posts), and test data (1,488 posts). The study utilized a range of advanced deep learning models, including GRU, Bi-GRU, CNN, LSTM, and BiLSTM, in addition to employing transfer learning techniques with BERT, XLNet, and RoBERTa. RoBERTa outperformed both XLNet and BERT with an f1-score of 0.83%, while XLNet and BERT earned scores of 0.80% apiece. Another study [32] examined Reddit social media data to identify and classify postings pertaining to mental health. The study identified eleven mental health topics, including BPD, bipolar disease, schizophrenia, anxiety, depression, self-harm, suicidewatch, addiction, severe alcoholism, opioids, autism, and non-mental health. Several models were utilized, including a feed-forward model with a precision of 70.82%, a convolutional neural network model with a precision of 71.37%, a linear model with a precision of 58.72%, and an SVM model with a precision of 64.02%.

This study requires the development of a predictive model for mental health categorizations. The dataset utilized in this study is derived from textual content sourced from the online platform Reddit. It encompasses information pertaining to the topics of depression and anxiety. The dataset encompasses seven distinct classification labels, namely borderline personality disorder (BPD), anxiety, depression, bipolar disorder, mental illness, schizophrenia, and poison. The model generation procedure utilizes a combination of Convolutional Neural Network and Bidirectional Long Short-Term Memory. CNN is employed to extract local features from the text, while BiLSTM offers the advantage of comprehending the contextual information of preceding and subsequent words through its forward and backward layers. The depth of the model's processing is increased, hence enhancing its ability to comprehend the contextual nuances of the text. In addition, the utilization of Fasttext word weighting is employed in tandem with CNN-BiLSTM to generate word representations for terms that are absent from the training data or to tackle out-of-vocabulary challenges. In the event that some words are absent or unattainable during the training phase, these words will be reconstructed into a series of syllables in order to acquire a vector representation [33]–[36].

2. Related Work

In the recent past, social media has emerged as a prominent platform for those seeking help with mental health concerns. This phenomenon stimulates researchers to extract the accessible data by utilizing natural language processing and machine learning methodologies, offering support to individuals requiring aid. While first study mostly concentrated on analyzing content on Twitter [37]–[39], subsequent interest shifted towards the Reddit platform [32], [39]–[41]. Text analysis pertaining to mental health has employed a diverse range of methodologies, which includes both machine learning techniques and advanced deep learning approaches. As an illustration, [37] employed character-level language models to evaluate the inclination of individuals with mental health issues to adopt a particular group of characters. The authors of the study [39] utilized multi-task neural learning (MTL), regression, and multi-layer perceptron single-task learning (STL) models to classify different forms of mental health illnesses. In a separate investigation [42], a support vector machine (SVM) model was trained to categorize 200 text messages into two distinct groups: ADHD or non-ADHD. An essential stage in the procedure was eliminating ADHD-associated keywords from the messages prior to learning, with the objective of assessing the SVM's ability to learn in the absence of pertinent keywords and semantic content.

The application of Deep Feed Neural Network, which has demonstrated superior performance in comparison to conventional machine learning models [43], [44], has been utilized in the analysis of clinical and genetic data to forecast mental health diseases. As an illustration, the diagnosis of depression included word embedding's in conjunction with neural networks like convolutional neural networks (CNN) and recurrent neural networks (RNN) [38]. In the domain of binary classification for mental health-related text posts, [32] employed Feed Forward Neural Networks, CNN, conventional machine learning techniques like SVM, and linear classification. A hierarchical attention network was trained to perform binary classification for each mental disorder, including depression, ADHD, anxiety, and others, resulting in [45] detections. A recent advancement entails the use of a CNN-based classification model [41], wherein distinct binary classifications are trained for each specific mental disease. In a broader sense, [46] investigated various factors that could impact mental health during the COVID-19 pandemic. They employed machine learning algorithms such as Naïve Bayes, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, and Gradient Boosting to analyze the data. Additionally, they examined a feature selection technique called LIME.

The study [22] employed a dual-method strategy, utilizing machine learning and deep learning techniques to detect and diagnose mental health issues in Reddit data. The machine learning techniques employed were LinearSVC, Logistic Regression, Naïve Bayes, and Random Forest. On the other hand, deep learning models encompassed GRU, Bi-GRU, CNN, LSTM, and Bi-LSTM, along with transfer learning methods utilizing BERT, XLNet, and RoBERTa. Despite the suboptimal results observed in previous research on deep learning algorithms like CNN, LSTM, and Bi-LSTM, this study seeks to enhance the accuracy of identifying mental health conditions. The approach involves creating a model that integrates Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) architectures. The model specifically focuses on recognizing seven mental health labels: borderline personality disorder (BPD), bipolar disorder, depression, anxiety, schizophrenia, mental illness, and poison. According to the findings in reference [47], utilizing CNN-BiLSTM for classifying depression based on Twitter data yields the most optimal outcomes.

3. Research Method

The research included several phases, as depicted in Figure 1 the data utilized in this study was obtained from textual sources pertaining to individuals experiencing anxiety and depression. Additionally, a separate corpus of textual data was gathered from user comments on the online platform Reddit. Upon concluding the data gathering procedure, the subsequent stage entailed executing pre-processing techniques. This encompassed the conversion of text to lowercase, removing numbers, punctuation marks, and stop words, alongside the process of tokenization. The dataset was subsequently divided into two subsets: the training dataset and the test dataset, Once the split stage has been finalized, the subsequent step involves the incorporation of those words into the word embedding process. In the present work, the fasttext algorithm [48] was employed to assign a numerical value or vector representation to individual words. Following the completion of the word weighting phase, the subsequent stage involved modeling the data by configuring the parameters of the CNN-BiLSTM, which encompassed the optimizer and activation function. During the assessment stage, the model's accuracy, precision, recall, and F1-Score were computed to assess its performance.

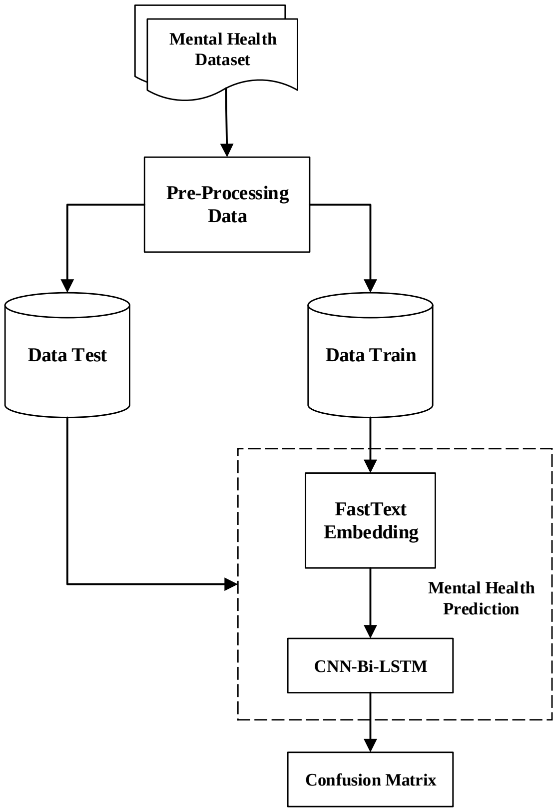

Figure 1. Proposed Mental Health CNN-BiLSTM Model

3.1 Data Collection

During the initial stage, data were collected on the subject of mental health. The dataset utilized in this study was retrieved from two different sources: the mental health text corpus and the dataset consisting of comments related to mental health concerns, such as anxiety and depression. It is organized into two columns: one for the comments themselves and another for toxicity labels indicating the level of harm. The data training procedure utilized 5,000 toxicity labels. Table 1 displays an example of the Mental Health Corpus, acquired from Kaggle (https://www.kaggle.com/datasets/reihanenamdari/mental-health-corpus/data) and the Reddit mental disorders identification dataset, which was also obtained from Kaggle (https://www.kaggle.com/datasets/kamaruladha/mental-disorders-identification-reddit-nlp). The Mental Health Corpus comprises five columns: title (the heading of the post), selftext (the body of the post), created_utc (the timestamp of the post's creation), and over_18 (the label indicating if the post contains NSFW content), and subreddit (the platform where the post was published). Table 2 presents a concrete example of the mental health dataset obtained from Reddit. The merged text was obtained by combining the selftext and title columns. The columns created_utc and over_18 were removed, and subreddit was utilized as a label. The Reddit dataset consists of 700,000 data points, with 5,000 data points chosen for each label throughout the training process. The data was subsequently compiled into an Excel spreadsheet to produce seven unique categorizations related to mental health, encompassing borderline personality disorder (BPD), anxiety, depression, bipolar disorder, mental illness, schizophrenia, and poison. The dataset was partitioned into two distinct subsets, specifically, the training data and the testing data. The training dataset comprised 35,000 entries, whereas the testing dataset comprised 6,108 entries.

Table 1. Mental Health Corpus Example

Text	Label
im done trying feel better the reason im still alive know mum devastated ever killed myself ever passes im still state im going hesitate ending life shortly after im almost take meds go therapy nothing seems help enough dont want around anymore hate feeling like this wouldnt wish upon enemy brain feels like constantly like static tv wont shut overthinking do think im running options dont see living past got accepted health science degree dont even know wanna try know im smart mental illness holds back think cant anything im good enough need fucking help dont know anymore ive run options	poison

want die even tho feels like dead years hello everyone want tell u depression loneliness general feeling wanting end life good dont even know im this idk im looking posting so im ashamed problems point feel ashamed writing strangers wont ever see life since like years ago problem bad breath really mean bad cant erase even directly brushed teeth trying heal sickness long time little success many doctors free health care dont live us obviously im unemployed cant afford pay doctor stopped trying go doctors one doctor laughed sickness literally could feel didnt care gave vibe its sickness go home imagine struggling even talk openly doctor receiving ur nuts sick treatment him im sure people dont problem hard understand hard me let put like would u feel u stepped shit couldnt wipe off would u comfortable walking around standing next someone talking someone obviously dont even need talk someone smell it breathing thru nose like result got anxiety standing someone going shop like climbing mount everest me achieve nothing life none want anything person smells none ever take serious someone like that want able interact people	poison
giving thoughts ive wanting od since im almost getting worse started wasnt bad easy distract thoughts pretty infrequent cant get head things much complicated thoughts mainly bother me part wants give make stop	poison

Table 2. Mental Health Reddit Dataset Example

Title	Selftext	Created_Utc	Over_18	Subreddit
does anyone else struggle with adderall crashes?	so ive been diagnosed with bpd, adhd and a million other things but these two are the ones that i struggle with the most. adderall helps me in ways that i cant even describe at times, its the only thing that makes me feel like a human. i take lamictal as a mood stabilizer as well, both extended release. the crash from the adderall though gets rough. thats when i start spiraling at times, my thoughts go crazy and i feel obsessive/incredibly depressed/sometimes start crying, you know the drill. sometimes it triggers splitting too. i feel like i somewhat have a handle on it? like rationalizing my thoughts, telling myself that iam not a monster and i feel the way i feel and thats valid, i just cant act on any crazy thoughts i have, stuff like that. maybe iam answering my own question here but does anyone have any advice for how to make the crashes more bearable? i know it gets better and it will wear off, i just hate going through it until it does	1650326982	FALSE	BPD
I've just recently been diagnosed with a bipolar disorder	I have really bad I guess mood switches. Certain things just cause me to flip the fuck out, literally the smallest things. When I start to have an episode, I start getting every emotion out at once and it can range from angry yelling to crying to laughter. I'm medicated (lamotrigine) but I still definitely will still get really pissed of at times. The mood swings haven't been happening as much, and when I'm angry I have been able to control it a little better. Does anyone have advice for dealing with something like this? Is there another medication hall suggest? How do you carry on your everyday life with this type of disorder Note: I do have other disorders like anxiety, depression, adhd.	1668481825	FALSE	bipolar
Should I just end it?	Honestly I'm fucking done with life over 1543 people have told me to kill myself I still	1662092365	TRUE	depression

	remember the first person who told me to kill myself. I hate my school they do jack shit about bullying I've been threatend to being stabbed 3 times and I can't take it anymore I wish I was dead so my school would be happy they all hate me.			
Started cutting, should I tell anyone?	I have no idea if this is nsfw or not. Thought I'd play it safe. I'm 14 and suffer from anxiety which has been especially hard over Christmas with the looming dread of school starting. Last night and the night before I was extremely anxious and wondered if cutting myself might offer some type of brief relief. I only cut deep enough to bleed on the tip of my thumb but scratched my arm and leg so it wasn't very bad but I found it very addictive, I'm now wondering if I should tell anyone about it. I have a meeting with my psychotherapist on Tuesday but I'm worried he might have to tell my parents and I really don't want that to happen as I'd feel extremely anxious about it. What should I do?	1640837671	TRUE	Anxiety
New Here	Hey everyone. Thanks for being a part of this group, and nice to meet you all. I hear voices sometimes. I want them to go away, and sometimes they don't - even on strong medicine. I want to date and be normal. How do you explain your schizophrenia? Is it a turn off? Thanks again.	1644005470	FALSE	schizophrenia
Why do I see self harm as enjoyable (I did stop due to increased surveillance but this question has been on my mind recently)	Well to put it simply, I recently started self-harming enough to leave scars on my body. Before I was kinda hesitant to feel pain and even when I held a knife so badly wanting to cut, I never could. But now I can cut myself pretty easily and started poking my arms with a needle very easily. Like I'm somewhat proud of my scars. But only slightly ashamed when my professors literally calling me out in front of the entire class about why I have bandages on (so I make up stupid lies). But other than that I kinda flaunt them to people around me. Like yeah I'm mentally unstable what about it	1668668048	TRUE	mentalillness

3.2 Pre-Processing

Pre-processing provides the next step after to the collection of datasets pertaining to mental health, preceding the commencement of the modeling process. Datasets pertaining to mental health necessitate undergoing multiple stages of data cleansing. These stages involve converting text to lowercase letters [49], eliminating non-classifying numbers to enhance efficiency [50], excluding punctuation marks such as periods, commas, question marks, exclamation marks, and semicolons [51], removing irrelevant words like conjunctions and prepositions [52], and performing tokenization to segment sentences into individual words.

3.3 Word Embedding

FastText is a word weighting technique that was derived from the Word2vec method[53]. The AI research team at Facebook derived influence from Mikolov's study [54] in the development of their model. The study presented evidence that FastText has the capability to effectively train a corpus of 1 billion words within a time frame of 10 minutes, while yet achieving comparable outcomes to other existing models [48]. The architectural design of the FastText model, akin to the Continuous Bag-of-Words (CBOW) approach employed in Word2Vec, has a hierarchical structure and encodes words as dense vectors. According to the findings presented in Figure 2, FastText incorporates a concealed layer positioned between the input and output layers [48].

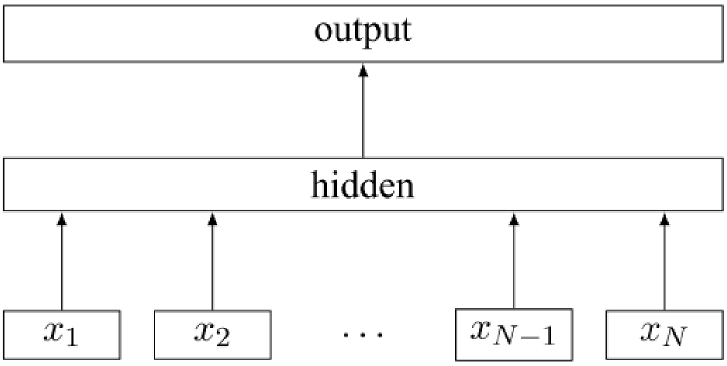

Figure 2. Architecture of the FastText Model for Sentences with N-gram Features [48]

The evaluation of word vectorization models usually overlooks the consideration of word morphology, consequently limiting the representation of terms in languages characterized by huge vocabularies and a substantial proportion of rarely employed words [33]. The difficulty is addressed by FastText through the provision of word representations for words that are absent in the training data, facilitating efficient model training on extensive datasets. In instances where a word is not encountered during the training of a model, the embedding vector can be computed by segmenting it into n-grams. The FastText model employed in this study consists of 300 dimensions and encompasses a collection of 2 million word vectors. These vectors were trained using subword information extracted from the Common Crawl https://fasttext.cc/docs/en/english-vectors.html.

3.4 Deep Learning Model

The process of determining hyperparameters is often conducted before to the evaluation of the dataset. After applying the fasttext word weighting technique, the textual input was converted into vector representations. These vectors, with a dimensionality of 300, were then passed into the embedding layer. The classification approach started with the input layers, namely Conv1D, MaxPool1D, and bidirectional LSTM. After the initial processing in the input layer, the data related to mental health was passed to the Con1D layer, which was utilized to control the intensity. Subsequently, the input representation underwent the MaxPool1D procedure, wherein the maximum value in the other dimensions was chosen to reduce the scale of the input. Finally, the CNN-BiLSTM model was subjected to the Bidirectional LSTM layer, which incorporated a dropout rate of 0.3 in order to mitigate the issue of overfitting during the training process. The results obtained from the computation process were communicated to the output layer. The output layer consisted of a dense layer utilizing the softmax activation function. Table 3 presents the parameter configurations employed for the modeling process utilizing the CNN-BiLSTM approach.

Table 3. Hyperparameter Settings used

Layer	Type	Neuron
Input Layer	Embedding	2662
	Dropout	0.3
Conv1D	Filter	64
	Kernel Size	3
MaxPool1D	Pool Size	3
	Dropout	0.3
Conv1D	Filter	64
	Kernel Size	3
MaxPool1D	Pool Size	3
	Dropout	0.3
LSTM Layer	Bidirectional LSTM	128
	Dropout	0.3
	Dense	128
	Dropout	0.3

Output Layer	Dense	7

3.4.1 CNN

A convolutional neural network (CNN) is a type of deep learning system that utilizes images as its input data. CNN utilizes the convolution technique to extract features from the input matrix, wherein a matrix operation is performed. In the context of text data processing, Convolutional Neural Networks, are employed to substitute pixels or images with character-based data. In this procedure, the textual data is seen as a matrix of input data.

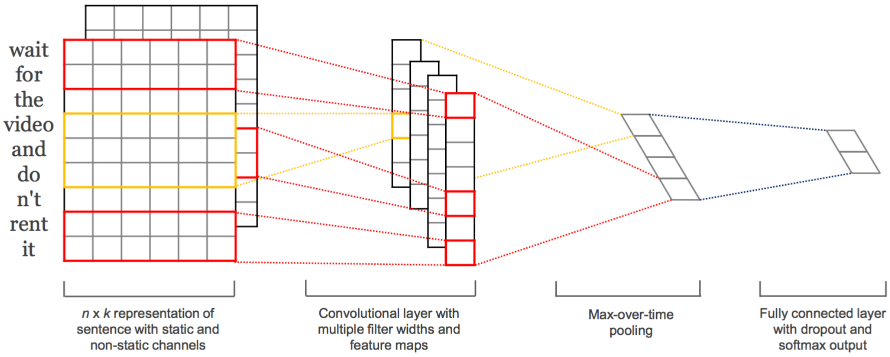

Figure 3. CNN Architecture

In this study, a convolutional neural network with one dimension (1D CNN) was utilized. The network architecture comprised an input layer, a convolutional layer, a max pooling layer, and a fully connected layer.

a. Input Layer
   The mental health datasets are transformed into a 300-dimensional word vector representation using FastText word embedding. This representation is subsequently utilized as a document vector.
b. Convolutional Layer
   The convolution layer consists of 64 filters, each with a window size of 3. These filters are selectively arranged in a pattern to get the desired filter outputs. The filters are applied vertically to the entire input matrix. The convolution action is performed by applying the dot product of the filter weights and the input matrix weights, followed by a non-linear operation using the Swish activation function. Ultimately, the filter weights are multiplied by the weights of the input matrix.
c. MaxPooling
   The objective of this layer is to optimize the value derived from the window size component by a factor of 3, so guaranteeing that the convolved feature map captures the most significant information.
d. Fully Connected
   The subsequent stage involves creating a connection among the feature map, which was produced in the preceding hidden layer, and the output layer for the purpose of classification. The layer in question employs the softmax activation function and loss function, and it encodes the multiclass output variable through the one-hot encoding technique.

3.4.2 Bi-LSTM

Progressive text categorization employs artificial neural networks, specifically recurrent neural networks with long short-term memory (LSTM) units. One limitation of Recurrent Neural Networks is their constrained storage capacity, which restricts their ability to retain and store information over extended periods of time. In order to address this limitation, researchers have developed the Long Short-Term Memory (LSTM) concept. The architecture of Bi-LSTM is depicted in Figure 4, illustrating its bidirectional nature and including two layers. This model demonstrates progress in comparison to the Long Short-Term Memory paradigm. In the present model, a sequential evaluation of each word is conducted, resulting in a heightened efficacy in the detection of patterns inside phrases. The forward layer of the model is responsible for sequentially processing words in the order they appear, whereas the backward layer processes words in the opposite direction, starting with the last word and moving towards the first word. The model is equipped with two bidirectional layers, which allows it to effectively capture the contextual information of both preceding and current words. This facilitates the model's full comprehension of the contextual information inside the text.

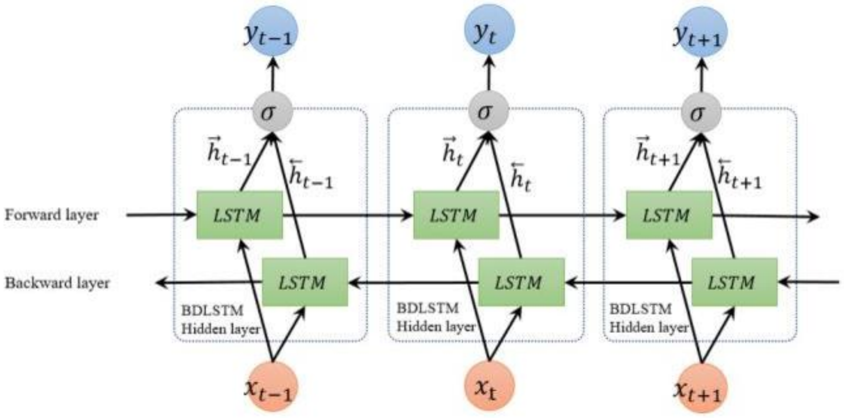

Figure 4. Bi-LSTM Architechture [55]

3.5 Evaluation Metrics

The evaluation of the model's performance was conducted by employing a confusion matrix on the testing dataset. The performance of this method was evaluated based on metrics such as Accuracy, Precision, Recall, and F-1 Score. The concept of accuracy pertains to the degree of proximity between the anticipated value and the actual value, as quantified by Equation 1. The term True Positive (TP) denotes instances when positive data is accurately predicted, while True Negative (TN) refers to instances where negative data is accurately predicted. The term False Positive (FP) refers to the situation when negative data is erroneously classified as positive data, whereas False Negative (FN) is used to describe the scenario where positive data is inaccurately classified as negative data.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Precision is a metric that evaluates the degree of accuracy in data by considering both the accuracy of the information and its ability to predict. Equation 2 is utilized for the purpose of articulating precision.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

The recall metric quantifies the effectiveness of the model in accurately identifying a specific class. The calculation can be performed utilizing Equation 3.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

The F1-Score is a metric that integrates two fundamental aspects of model evaluation, namely precision and recall. Precision is a metric that quantifies the degree of accuracy exhibited by a model in predicting data. On the other hand, recall is a metric that evaluates the effectiveness of a model in correctly identifying instances belonging to a particular class. The F1-Score offers a comprehensive assessment of the model's predictive accuracy by integrating multiple metrics. The F1-Score is computed using Equation 4.

$$F1 - Score = 2\,\frac{(Recall \times Precision)}{(Recall + Precision)} \tag{4}$$

4. Results and Discussion

4.1 Dataset

The dataset pertaining to mental health has been categorized into seven distinct labels. Borderline Personality Disorder (BPD) is a psychiatric disorder characterized by impulsive behavior, emotional instability, unstable interpersonal relationships, and a distorted sense of self [56]. Anxiety is a mental health issue characterized by excessive anxiety, worry, and behavioral disturbances [57]. Depression is a psychological condition characterized by feelings of sadness, loss of interest in daily activities, decreased appetite, and difficulty focusing [58]. Bipolar disorder, previously referred to as manic depressive illness, is characterized by abnormal fluctuations in mood, energy levels, activity, and concentration [59]. Mental illness is a condition that impacts an individual's thoughts, mood, behavior, and cognitive functioning [60]. Schizophrenia is a psychiatric disorder that affects an individual's thoughts, emotions, and behavior, and is characterized by hallucinations and delusions [61]. The text additionally addresses the topic of poison, which is commonly linked to individuals who are afflicted with symptoms of anxiety, despair, and other mental health disorders. During the process of classifying the dataset, the label "depression" is assigned by using the keyword "depression," whereas the category "schizophrenia" is determined by identifying keywords such as "hallucinating" and "schizophrenia." The use of the phrases "mental health" and "mental illness" [62], [63] suggests the designation of "mental illness." In order to detect borderline personality disorder, also known as BPD, it is necessary to focus on certain phrases such as "BPD" and "borderline personality disorder" [64], [65]. Bipolar disorder can be identified by specific phrases such as "bipolar" [66]. The term "anxiety" is associated with the terms "anxiety" [67], [68]. Furthermore, the term "depression" is linked to terms such as "depression" and "anxiety" [62], [67].

Table 4. Mental Health Text Classification Sample Data and Labels

Text	Label
What should I do? I have a behavioral health assessment in the beginning of next month. For perspective, I've come to the realization that I have BPD, but because of it, I keep thinking that all the times Im spiraling are just me overreacting. I am 17, so I know professionals are hesitant to diagnose under 18. I kind of dont want to get admitted when the school year is ending and I have a bunch of tests to attend, but I need to get my shit together before Im 18 because I honestly dont think I can make it. If Im honest about everything, I most likely get admitted. I just dont know if its the right move. Is it going to be worth it?	BPD
Does anyone deal with no support from family? I was diagnosed with bi polar2 and OCD a little over a year ago. Unfortunately I had a traumatic event at my wedding that seemed to have triggered both. I have never felt very supported by my family. In fact whenever I try to talk to them about my OCD or Bi Polar they act annoyed, like its not real and Im just being dramatic or needy. My husband is extremely supportive though and is constantly doing research to understand how to best help me so I am very lucky in that sense I guess I was just wondering if anyone else feels almost forgotten about or brushed off due to their Bipolar disorder, how did you handle it?	bipolar
I am so tired of being depressed since almost 10 yrs while others around me lead normal lives I've probably been depressed since I was 12 (I'm 22 now) and it just keeps getting worse and worse. If I was made aware of how horrible my mental health would become in the future I would've killed myself long ago. Over time, I found a decent support system with my friends who also suffered from their fair share of problems and mental health issues but over the past 2-3 months things have really been terrible for me. I can see everyone else move on with their lives, enjoy good times, and feel good stuff. Meanwhile I keep having terrible mood swings and can only feel good if I throw myself into a stupid false reality. My depressive phases and rock bottom moments have become more and more common and I feel like a side character in my own story. I feel like I have no support system at all and tbh, I'm okay being all by myself but seeing others be normal while I suffer is damn painful.	depression
Anyone else get random anxiety for just talking to someone either on the phone or person? Could be anyone, friends, family, work colleagues. When i start talking i start feeling shaky, my voice goes, lightheaded and a little dizzy. Its light something triggers when i start talking no matter what. I know sometimes i talk and prob dont breath properly when talking fast by still. If I talk longer than 10 mins me head feels woozy like i have had one too many.	anxiety
I'm Hallucinating, help. I keep seeing things in the corner of my eyes and whenever I close my eyes. I hear my breathing playback... This disease is weird. Any help or anything I can't think please help thanks it's appreciated	schizophrenia
Mental health I don't know who to talk to about mental illness and I think my mental illness is just getting worse. I never talk to ANYONE about this because I'm scared they would just see me as attention seeking. *note* I'm not diagnosed with any mental illness and aren't asking a diagnosis. But I'm just so scared of being talked about behind my back. I keep on yelling at people for no reason. I'm kind of sad most of the time. So yeah...	mentalillness

dont belong anywhere ive tried years fit depression anxiety feels like im fundamentally different peers try fit people always end fucking never happy dont belong here one energy put depressing bullshit drag people make life worse im gonna make past highschool hate much deserve rot	Poison

4.2 Pre-Processing Result

Figure 5 illustrates the results of the data cleansing procedure conducted in the pre-processing stage of the mental health dataset. The procedure entails the conversion of text to lowercase, the elimination of digits, punctuation marks, and stopwords, and the subsequent tokenization.

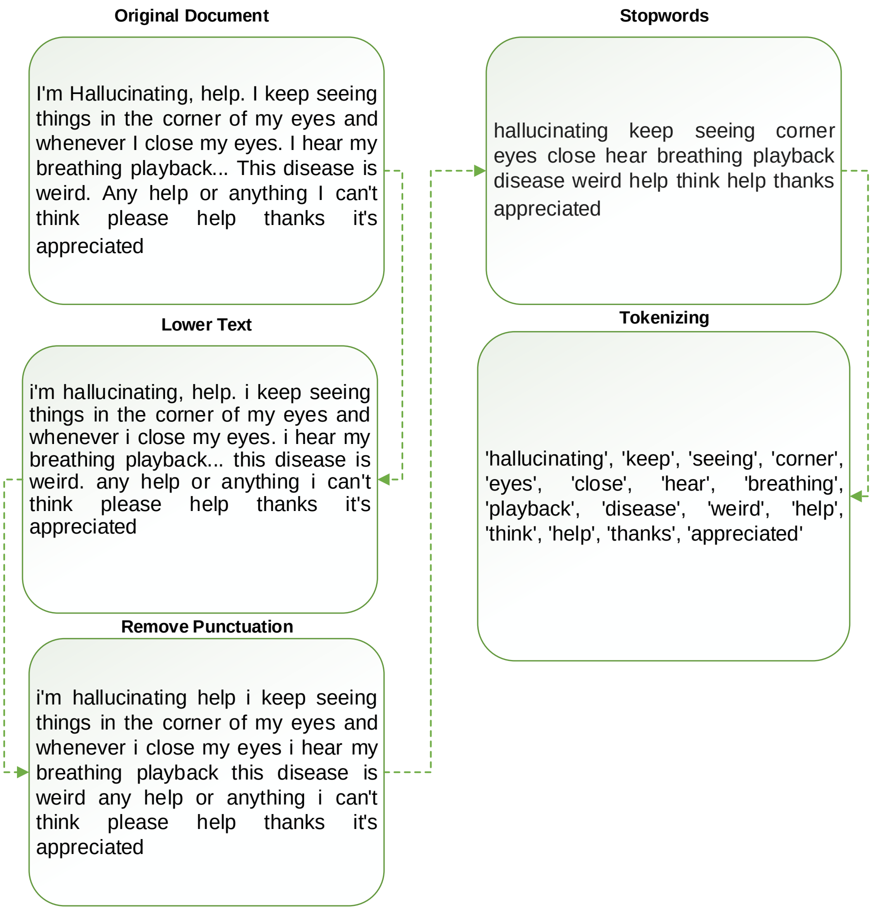

Figure 5. Pre-Processing Result

4.3 Hyperparameter Setting

Before starting the review process of test results, the mental health dataset were subjected to training utilizing the training approach, employing 21 epochs and a batch size of 64. The dataset employed the RMSprop optimizer for the purpose of training. The training accuracy results for the CNN-BiLSTM and BiLSTM models can be found in Table 5.

Table 5. Model Training Data Results

Model Algorithm	Optimizer	Accuracy
CNN-BiLSTM	RMSprop	0.89%
BiLSTM		0.88%

Each scenario was subjected to testing using a confusion matrix that utilized the classification methodology. After the completion of the training procedure utilizing the mental health dataset, the subsequent phase involved evaluating the accuracy and f1-score of the classification findings. The CNN-BiLSTM deep learning model outperformed the BiLSTM model, as evidenced by the results presented in Table 6. Both models demonstrated satisfactory performance in addressing multi-class classification tasks. Although there was a minimal disparity in the performance of the BiLSTM and CNN-BiLSTM models, the CNN-BiLSTM algorithm earned the highest overall accuracy score of 0.85%, while the BiLSTM model achieved an accuracy value of 0.83%.

According to this study, CNN-BiLSTM outperforms other deep learning models, such as CNN, BiLSTM, or LSTM, in accurately detecting mental disorders. Research is crucial in improving the precision of the model, especially in comparison to prior studies. For example, a previous study that used LSTM as a classification model obtained an accuracy of only 0.76% for post titles [21], while another study reported an even lower accuracy rate [22]. The investigation revealed that the CNN model attained an accuracy of 0.64%, whilst the LSTM obtained 0.76%, and the BiLSTM outperformed both with an accuracy of 0.78%. Our suggestion recommends employing 7 classification categories: BPD, bipolar disorder, depression, anxiety, schizophrenia, mental disease, and poison. In contrast to the study [22], which employed only six categories (adhd, anxiety, bipolar, depression, ptsd, and none), this research takes a different approach. Previously, a study [21] categorized individuals into six distinct classes for classification purposes: ADHD, anxiety, bipolar disorder, depression, PTSD, and none. According to our study presented in Table 6, the use of CNN-BiLSTM model demonstrates a significant competitive advantage when compared to similar studies. In addition, this study offers convincing evidence of improved accuracy and f1-score. Our proposed model does not only incorporate a larger number of classes, but it also produces improved outcomes, consequently showcasing its effectiveness in categorizing a diverse array of mental health disorders.

These findings make an important contribution to comprehension and implementation of deep learning models for the classification of mental disorder. They provide evidence that the CNN-BiLSTM technique exhibits greater potential to enhance accuracy in this domain. The empirical findings indicate that the utilization of FastText weighting yields superior accuracy values for the classification of mental health datasets with seven labels, for both the CNN-BiLSTM and BiLSTM models. Table 6 displays the observed enhancements in accuracy resulting from the utilization of the two distinct approaches.

Table 6. Examining the Accuracy Comparison among Deep Learning Models

Model Algorithm	Class	Accuracy	F1-Score
CNN-BiLSTM	bpd, bipolar, depression, anxiety, schizophrenia, mentalillness, poison	0.85%	0.85%
BilSTM	bpd, bipolar, depression, anxiety, schizophrenia, mentalillness, poison	0.83%	0.83%
CNN [22]	adha, anxiety, bipolar, depression, ptsd, none	0.64%	0.65%
LSTM [22]	adha, anxiety, bipolar, depression, ptsd, none	0.76%	0.77%
BiLSTM [22]	adha, anxiety, bipolar, depression, ptsd, none	0.78%	0.79%
LSTM [21]	adhd, anxiety, bipolar, depression, ptsd, none	0.76%	0.76%

The results of classification obtained from the confusion matrix utilizing the CNN-BiLSTM model are presented in Table 7. The surprise outcome of the study is the model's capacity to effectively identify patterns associated with toxic content pertaining to mental health. The model demonstrates a high level of accuracy in classifying the poison class, achieving an F1-Score of 0.98%. This finding indicates that the model exhibits a significantly low percentage of false positives in identifying mental disorders through the analysis of social media data.

According to the F1-Score metric, the classes that demonstrate the highest performance are anxiety 0.90%, borderline personality disorder 0.87%, depression 0.82%, and schizophrenia 0.81%. In contrast, the two classes with the lowest performance rates are bipolar disorder 0.79% and mental illness 0.70%.

Table 7. Results of the Confusion Matrix for Each Class in CNN-BiLSTM Model

Class	Precision	Recall	F1-Score
Anxiety	0.89%	0.90%	0.90%
BPD	0.90%	0.84%	0.87%
Poison	0.99%	0.98%	0.98%
bipolar	0.86%	0.73%	0.79%
depression	0.78%	0.86%	0.82%
mentalillness	0.73%	0.67%	0.70%
schizophrenia	0.75%	0.87%	0.81%

The results of predicting the confusion matrix for each label using the CNN-BiLSTM model are depicted in Figure 6. The classification label 'poison' has a notable level of accuracy, as it predicts accurately 964 out of 988 text entries. The algorithm accurately identified the anxiety label in 902 instances of text data, followed by schizophrenia 776 examples, depression 751 examples, and BPD 721 examples. Nevertheless, the model exhibited a level of accuracy in predicting the bipolar label by accurately classifying 570 out of the total 779 text elements. The classification outcomes pertaining to mental illness were influenced by the inclusion of texts that encompassed a broad scope of mental health topics. During the training sessions, there was a noticeable inclination towards prioritizing other classes in the materials. Table 8 of the document includes many terminologies, including BPD, bipolar disorder, anxiety disorder, and depression, in addition to the overarching concept of mental health. The terminology used is characterized by a lack of clarity or multiple interpretations within the scope of this study, the model is limited to dealing with extremely lengthy sentences. This limitation may result in a biased classification of sophisticated language structures and contexts in specific datasets [24]. There are several potential interpretations. According to the analysis of the confusion matrix, it was found that a total of 714 data points were examined. Among these, only 475 data points were correctly classified under the mental disease category. The model is a first step towards practical application of mental health detection. This study demonstrated the model's innovative use as an early warning system, providing valuable assessments for professionals in the monitoring of mental health patients.

Table 8. Another Class with Text and Mental Health Labels that Lean Towards Mentalillness

Text	Label
not know anymore lose not sure expect post figure could not hurt first time poster way struggle mental health issue pretty much whole life diagnose run gammut bpd bipolar anxiety depression ocd ptsd hypervigilance abandonment disorder addition issue physicalhormonalneurological tell close year qualify disability finally apply route november see state doctor deny because physically capable accommodation appeal see state judge seem convinced file stay home mom not case despite employment expert state unable gainfully employ base mental neurological issue deny appeal decision via internet turn friday file civil appeal request extension know not wait finally request referral begin mh counselling since maybe med check pcp year time pcps psychiatrist counselor due practitioner leave practice insurance change not able afford care mentally could not start someone new tired struggle tired feeling worthless like lazy pos not know healthy damn expensive jump three ring circus prove mental illness treat year without much progress hate rob peter pay paul decide buy grocery pay bill take care health care tired look like husband make much money get kind help bebecause not consider mortgagebillsme care worry kid lunch account might get turn bebecause far red contact ssi law firm would not take case bebecause nothing change bebecause not see mental health care provider happy expound info herein cliff note version crap not event know	mentalillness

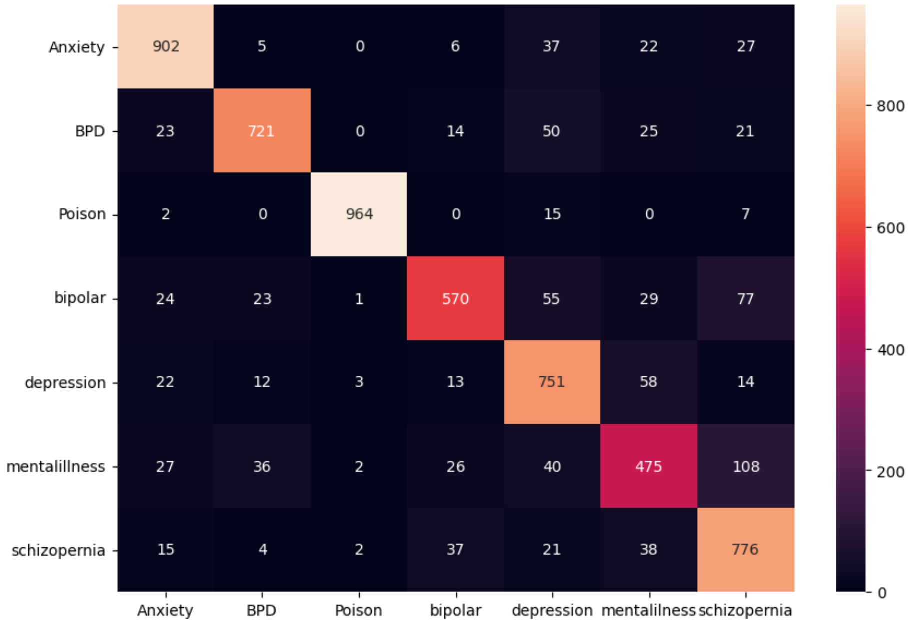

Figure 6. Confusion Matrix Result

4. Conclusion

Over the past few decades, social media has emerged as a prominent medium for individuals to convey their emotional states through several modalities such as textual content, pictures, and videos. The COVID-19 pandemic has resulted in a notable rise in mental health concerns and heightened vulnerability among those with pre-existing mental health conditions, owing to the extensive implementation of isolation measures. Consequently, there is a necessity for the development of a detection methodology that may furnish timely alerts and facilitate the diagnosis of individuals afflicted with mental health disorders. In order to tackle these concerns, the present study utilized the CNN-BiLSTM model as a prospective resolution. The model utilized text data pertaining to mental health, specifically tagged with terms such as Borderline Personality Disorder (BPD), anxiety, depression, bipolar disorder, mental illness, schizophrenia, and self-harm. The utilization of Fasttext's pre-trained word weighting is incorporated in this model. The evaluation of the model's categorization procedure is presented in Table 4. The achieved accuracy of the model was 0.85%, accompanied by an F1-Score of 0.85%. This performance surpassed that of the Bi-LSTM model, which exhibited an accuracy of merely 0.83% and an F1-Score of 0.83%. During the modeling process, the outcomes demonstrated that the model effectively identified patterns associated with mental health, resulting in substantial enhancements in accuracy and f1-score compared to previous research. Nevertheless, a challenge that arises is the model's struggle in comprehending the context of excessively lengthy sentences, which may lead to biased interpretations of sentence substance.

Acknowledgement

This research is funded by Master Thesis Research grant, no. 181/E5/PG.02.00.PL/2023, from Directorate of Technology Research and Community Service, Ministry of Education, Culture, Research and Technology, Indonesia.

References

[1] J. Rehm and K. D. Shield, “Global Burden of Disease and the Impact of Mental and Addictive Disorders,” *Curr. Psychiatry Rep.*, vol. 21, no. 2, pp. 1–7, 2019. https://doi.org/10.1007/s11920-019-0997-0

[2] Rokom, “Ministry of Health Reveals Mental Health Issues in Indonesia,” 2021.

[3] J. Singh, M. Wazid, D. P. Singh, and S. Pundir, “An embedded LSTM based scheme for depression detection and analysis,” *Procedia Comput. Sci.*, vol. 215, pp. 166–175, 2022. https://doi.org/10.1016/j.procs.2022.12.019

[4] T. Zhang, A. M. Schoene, S. Ji, and S. Ananiadou, “Natural language processing applied to mental illness detection: a narrative review,” *npj Digit. Med.*, vol. 5, no. 1, pp. 1–13, 2022. https://doi.org/10.1038/s41746-022-00589-7

[5] G. Gonzalez-Hernandez, A. Sarker, K. O'Connor, and G. Savova, “Capturing the Patient's Perspective: a Review of Advances in Natural Language Processing of Health-Related Text.,” *Yearb. Med. Inform.*, vol. 26, no. 1, pp. 214–227, Aug. 2017. https://doi.org/10.15265/iy-2017-029

[6] O. G. Iroju and J. O. Olaleke, “A Systematic Review of Natural Language Processing in Healthcare,” *Int. J. Inf. Technol. Comput. Sci.*, vol. 7, no. 8, pp. 44–50, 2015. https://doi.org/10.5815/ijitcs.2015.08.07

[7] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, “Natural language processing: An introduction,” *J. Am. Med. Informatics Assoc.*, vol. 18, no. 5, pp. 544–551, 2011. https://doi.org/10.1136/amiajnl-2011-000464

[8] J. Ive *et al.*, "Generation and evaluation of artificial mental health records for Natural Language Processing," *npj Digit. Med.*, vol. 3, no. 1, pp. 1–9, 2020. https://doi.org/10.1038/s41746-020-0267-x

[9] A. S. M. Venigalla, S. Chimalakonda, and D. Vagavolu, "Mood of India during Covid-19 - An interactive web portal based on emotion analysis of twitter data," *Proc. ACM Conf. Comput. Support. Coop. Work. CSCW*, pp. 65–68, 2020. https://doi.org/10.1145/3406865.3418567

[10] E. D'Avanzo, G. Pilato, and M. Lytras, "Using Twitter sentiment and emotions analysis of Google Trends for decisions making," *Program*, vol. 51, no. 3, pp. 322–350, Jan. 2017. https://doi.org/10.1108/PROG-02-2016-0015

[11] G. Castillo-sánchez, M. Franco-martín, G. Marques, E. Dorronzoro, I. De Torre-díez, and M. Franco-martín, "Suicide Risk Assessment Using Machine Learning and Social Networks : a Scoping Review," *J. Med. Syst.*, 2020. https://doi.org/10.1007/s10916-020-01669-5

[12] M. A. Franco-Martín, J. L. Muñoz-Sánchez, B. Sainz-de-Abajo, G. Castillo-Sánchez, S. Hamrioui, and I. de la Torre-Díez, "A Systematic Literature Review of Technologies for Suicidal Behavior Prevention," *J. Med. Syst.*, vol. 42, no. 4, 2018. https://doi.org/10.1007/s10916-018-0926-5

[13] S. Ji, S. Pan, X. Li, E. Cambria, G. Long, and Z. Huang, "Suicidal Ideation Detection: A Review of Machine Learning Methods and Applications," *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 1, pp. 214–226, 2021. https://doi.org/10.1109/TCSS.2020.3021467

[14] F. T. Giuntini, M. T. Cazzolato, M. de J. D. dos Reis, A. T. Campbell, A. J. M. Traina, and J. Ueyama, "A review on recognizing depression in social networks: challenges and opportunities," *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 11, pp. 4713–4729, 2020. https://doi.org/10.1007/s12652-020-01726-4

[15] N. Mahdy, D. A. Magdi, A. Dahroug, and M. A. Rizka, "Comparative Study: Different Techniques to Detect Depression Using Social Media," 2020.

[16] A. Khan, M. S. Husain, and A. Khan, "Analysis of Mental State of Users Using Social Media To Predict Depression! A Survey," *Int. J. Adv. Res. Comput. Sci.*, vol. 9, pp. 100–106, 2018.

[17] S. Chancellor and M. De Choudhury, "Methods in predictive techniques for mental health status on social media: a critical review," *npj Digit. Med.*, vol. 3, no. 1, 2020. https://doi.org/10.1038/s41746-020-0233-7

[18] E. A. Ríssola, D. E. Losada, and F. Crestani, "A survey of computational methods for online mental state assessment on social media," *ACM Trans. Comput. Healthc.*, vol. 2, no. 2, 2021. https://doi.org/10.1145/3437259

[19] G. Coppersmith, M. Dredze, and C. Harman, "Quantifying Mental Health Signals in Twitter," in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Jun. 2014, pp. 51–60. https://doi.org/10.3115/v1/W14-3207

[20] R. A. CALVO, D. N. MILNE, M. S. HUSSAIN, and H. CHRISTENSEN, "Natural language processing in mental health applications using non-clinical texts," *Nat. Lang. Eng.*, vol. 23, no. 5, pp. 649–685, 2017. https://doi.org/10.1017/S1351324916000383

[21] A. Murarka and I. B. M. Raleigh, "Classification of mental illnesses on social media using RoBERTa," *Proc. ofthe 12th Int. Work. Heal. Text Min. Inf. Anal.*, pp. 59–68, 2021.

[22] I. Ameer, M. Arif, G. Sidorov, H. Gòmez-Adorno, and A. Gelbukh, "Mental Illness Classification on Social Media Texts using Deep Learning and Transfer Learning," 2022. https://doi.org/10.48550/arXiv.2207.01012

[23] W. Yue and L. Li, "Sentiment analysis using word2vec-cnn-bilstm classification," *2020 7th Int. Conf. Soc. Netw. Anal. Manag. Secur. SNAMS 2020*, pp. 3–7, 2020. https://doi.org/10.1109/SNAMS52053.2020.9336549

[24] M. Rhanoui and M. Mikram, "A CNN-BiLSTM Model for Document-Level Sentiment Analysis," pp. 832–847, 2019. https://doi.org/10.3390/make1030048

[25] L. Xiaoyan, R. C. Raga, and S. Xuemei, "GloVe-CNN-BiLSTM Model for Sentiment Analysis on Text Reviews," *J. Sensors*, vol. 2022, 2022. https://doi.org/10.1155/2022/7212366

[26] V. Tejaswini, K. S. Babu, and B. Sahoo, "Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, 2022. https://doi.org/10.1145/3569580

[27] Y. Kim, "Convolutional Neural Networks for Sentence Classification," 2014. https://doi.org/10.48550/arXiv.1408.5882

[28] K. Dheeraj and T. Ramakrishnudu, "Negative emotions detection on online mental-health related patients texts using the deep learning with MHA-BCNN model," *Expert Syst. Appl.*, vol. 182, no. May, p. 115265, 2021. https://doi.org/10.1016/j.eswa.2021.115265

[29] K. Zeberga, M. Attique, B. Shah, F. Ali, Y. Z. Jembre, and T. S. Chung, "A Novel Text Mining Approach for Mental Health Prediction Using Bi-LSTM and BERT Model," *Comput. Intell. Neurosci.*, vol. 2022, 2022. https://doi.org/10.1155/2022/7893775

[30] M. Trotzek, S. Koitka, and C. M. Friedrich, "Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 3, pp. 588–601, 2020. https://doi.org/10.1109/TKDE.2018.2885515

[31] S. Ghosal and A. Jain, "Depression and Suicide Risk Detection on Social Media using fastText Embedding and XGBoost Classifier," *Procedia Comput. Sci.*, vol. 218, pp. 1631–1639, 2023. https://doi.org/10.1016/j.procs.2023.01.141

[32] G. Gkotsis *et al.*, "Characterisation of mental health conditions in social media using Informed Deep Learning," *Sci. Rep.*, vol. 7, pp. 1–10, 2017. https://doi.org/10.1038/srep45141

[33] E. M. Dharma, F. L. Gaol, H. L. H. S. Warnars, and B. Soewito, "The Accuracy Comparison Among WORD2VEC, Glove, and Fasttext Towards Convolution Neural Network (CNN) Text Classification," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 2, pp. 349–359, 2022.

[34] M. R. Hossain, M. M. Hoque, and I. H. Sarker, "Text Classification Using Convolution Neural Networks with FastText Embedding," *Adv. Intell. Syst. Comput.*, vol. 1375 AIST, no. June, pp. 103–113, 2021. https://doi.org/10.1007/978-3-030-73050-5_11

[35] N. A. Hasanah, N. Suciati, and D. Purwitasari, "Identifying degree-of-concern on covid-19 topics with text classification of twitters," *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 7, no. 1, pp. 50–62, 2021. https://doi.org/10.26594/register.v7i1.2234

[36] T. T. Mengistie and D. Kumar, "Deep Learning Based Sentiment Analysis On COVID-19 Public Reviews," in *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2021, pp. 444–449. https://doi.org/10.1109/ICAIIC51459.2021.9415191

[37] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell, "CLPsych 2015 Shared Task: Depression and PTSD on Twitter," in *2nd Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, CLPsych 2015 - Proceedings of the Workshop*, 2015, pp. 31–39. https://doi.org/10.3115/v1/W15-1204

[38] D. I. Ahmed Husseini Orabi, Prasadith Buddhitha, Mahmoud Husseini Orabi, "Deep Learning for Depression Detection of Twitter Users," *Proc. ofthe Fifth Work. Comput. Linguist. Clin. Psychol. From Keyboard to Clin.*, pp. 88–97. https://doi.org/10.18653/v1/W18-0609

[39] A. Benton, M. Mitchell, and D. Hovy, "Multi-Task Learning for Mental Health using Social Media Text," 2017. https://doi.org/10.48550/arXiv.1712.03538

[40] A. Zirikly, P. Resnik, ¨ Ozlem Uzuner, and K. Hollingshead, "CLPsych 2019 Shared Task: Predicting the Degree of Suicide Risk in Reddit Posts," *Proc. Sixth Work. Comput. Linguist. Clin. Psychol.*, pp. 24–33, 2019. https://doi.org/10.18653/v1/W19-3003

[41] J. Kim, J. Lee, E. Park, and J. Han, "A deep learning model for detecting mental illness from user content on social media," *Sci. Rep.*, vol. 10, no. 1, pp. 1–6, 2020. https://doi.org/10.1038/s41598-020-68764-y

[42] M. Fekihal, J. Diederich, and A.-A. A, *Machine learning, text classification and mental health*. 2004.

[43] M. Amjad, N. Ashraf, A. Zhila, G. Sidorov, A. Zubiaga, and A. Gelbukh, "Threatening Language Detection and Target Identification in Urdu Tweets," *IEEE Access*, vol. 9, pp. 128302–128313, 2021. https://doi.org/10.1109/ACCESS.2021.3112500

[44] M. Amjad, G. Sidorov, A. Zhila, H. Gómez-Adorno, I. Voronkov, and A. Gelbukh, "'Bend the truth': Benchmark dataset for fake news detection in Urdu language and its evaluation," *J. Intell. Fuzzy Syst.*, vol. 39, pp. 2457–2469, 2020. https://doi.org/10.3233/JIFS-179905
[45] I. Sekulic and M. Strube, "Adapting deep learning methods for mental health prediction on social media," *W-NUT@EMNLP 2019 - 5th Work. Noisy User-Generated Text, Proc.*, pp. 322–327, 2019. https://doi.org/10.18653/v1/D19-5542
[46] Y. Hu and M. Sokolova, "Explainable Multi-class Classification of the CAMH COVID-19 Mental Health Data," no. December 2020, pp. 1–22, 2021. https://doi.org/10.48550/arXiv.2105.13430
[47] H. Kour and M. K. Gupta, *An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM*, vol. 81, no. 17. Multimedia Tools and Applications, 2022. https://doi.org/10.1007/s11042-022-12648-y
[48] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *15th Conf. Eur. Chapter Assoc. Comput. Linguist. EACL 2017 - Proc. Conf.*, vol. 2, pp. 427–431, 2017. https://doi.org/10.18653/v1/e17-2068
[49] M. R. Pribadi, D. Manongga, H. D. Purnomo, I. Setyawan, and Hendry, "Sentiment Analysis of the PeduliLindungi on Google Play using the Random Forest Algorithm with SMOTE," *2022 Int. Semin. Intell. Technol. Its Appl. Adv. Innov. Electr. Syst. Humanit. ISITIA 2022 - Proceeding*, no. July, pp. 115–119, 2022. https://doi.org/10.1109/ISITIA56226.2022.9855372
[50] M. Khader, A. Awajan, and G. Al-Naymat, "The Effects of Natural Language Processing on Big Data Analysis: Sentiment Analysis Case Study," in *2018 International Arab Conference on Information Technology (ACIT)*, 2018, pp. 1–7. https://doi.org/10.1109/ACIT.2018.8672697
[51] A. Squicciarini, A. Tapia, and S. Stehle, "Sentiment analysis during Hurricane Sandy in emergency response," *Int. J. Disaster Risk Reduct.*, vol. 21, no. December 2016, pp. 213–222, 2017. https://doi.org/10.1016/j.ijdrr.2016.12.011
[52] S. Sarica and J. Luo, "Stopwords in technical language processing," *PLoS One*, vol. 16, no. 8 August, pp. 1–13, 2021. https://doi.org/10.1371/journal.pone.0254937
[53] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," vol. 5, pp. 135–146, 2017.
[54] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *1st Int. Conf. Learn. Represent. ICLR 2013 - Work. Track Proc.*, pp. 1–12, 2013.
[55] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction," pp. 1–11, 2018. https://doi.org/10.48550/arXiv.1801.02143
[56] A. W. Bateman and R. Krawitz, "Borderline Personality Disorder: An evidence-based guide for generalist mental health professionals." Oxford University Press, May 01, 2013. https://doi.org/10.1093/med:psych/9780199644209.001.0001
[57] B. Storer *et al.*, "Global prevalence of anxiety in adult cardiology outpatients : A systematic review and meta-analysis The Black Dog Institute , Sydney , Australia School of Psychology , Faculty of Science , University of New South Wales , Sydney ," *Curr. Probl. Cardiol.*, p. 101877, 2023. https://doi.org/10.1016/j.cpcardiol.2023.101877
[58] A. H. Miller and C. L. Raison, "The role of inflammation in depression: from evolutionary imperative to modern treatment target," *Nat. Rev. Immunol.*, vol. 16, no. 1, pp. 22–34, 2016. https://doi.org/10.1038/nri.2015.5
[59] "Bipolar disorder," *Clin. Pediatr. (Phila).* https://doi.org/10.1177/0009922808316663
[60] 65 World Health Assembly, "Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level: report by the Secretariat." World Health Organization, Geneva PP - Geneva.
[61] E. Parellada and P. Gassó, "Glutamate and microglia activation as a driver of dendritic apoptosis: a core pathophysiological mechanism to understand schizophrenia.," *Transl. Psychiatry*, vol. 11, no. 1, p. 271, May 2021. https://doi.org/10.1038/s41398-021-01385-9
[62] I. Calixto, V. Yaneva, and R. M. Cardoso, "Natural Language Processing for Mental Disorders: An Overview," *Nat. Lang. Process. Healthc.*, no. October, pp. 37–59, 2022. http://dx.doi.org/https://doi.org/10.1201/9781003138013
[63] H. Herdiansyah, R. Roestam, R. Kuhon, and A. S. Santoso, "Their post tell the truth: Detecting social media users mental health issues with sentiment analysis," *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 691–697, 2022. https://doi.org/10.1016/j.procs.2022.12.185
[64] M. Lyons, N. D. Aksayli, and G. Brewer, "Mental distress and language use: Linguistic analysis of discussion forum posts," *Comput. Human Behav.*, vol. 87, no. May, pp. 207–211, 2018. https://doi.org/10.1016/j.chb.2018.05.035
[65] H. Dyson and L. Gorvin, "How Is a Label of Borderline Personality Disorder Constructed on Twitter: A Critical Discourse Analysis," *Issues Ment. Health Nurs.*, vol. 38, no. 10, pp. 780–790, 2017. https://doi.org/10.1080/01612840.2017.1354105
[66] E. Kadkhoda, M. Khorasani, F. Pourgholamali, M. Kahani, and A. R. Ardani, "Bipolar disorder detection over social media," *Informatics Med. Unlocked*, vol. 32, no. August, p. 101042, 2022. https://doi.org/10.1016/j.imu.2022.101042
[67] D. Levanti *et al.*, "Depression and Anxiety on Twitter During the COVID-19 Stay-At-Home Period in 7 Major U.S. Cities," *AJPM Focus*, vol. 2, no. 1, p. 100062, 2023. https://doi.org/10.1016/j.focus.2022.100062
[68] D. Zarate, M. Ball, M. Prokofieva, V. Kostakos, and V. Stavropoulos, "Identifying self-disclosed anxiety on Twitter: A natural language processing approach," *Psychiatry Res.*, vol. 330, p. 115579, 2023. https://doi.org/10.1016/j.psychres.2023.115579